A Cross-Country Analysis of the Determinants of Life Expectancy

Introduction to statsitics and econonometrics (L1022)

Table of Contents

[1. Introduction 3](#_Toc203923323)

[2. Descriptive Summary Statistics 3](#_Toc203923324)

[2.1 The Female Life Expectancy (LEF) 3](#_Toc203923325)

[2.2 The Male Life Expectancy (MEF) 4](#_Toc203923326)

[2.3 Gross Domestic Product Per Capita (GDPPC) 5](#_Toc203923327)

[2.4 Water Access (WA) 6](#_Toc203923328)

[2.5 Health Spending (HS) 6](#_Toc203923329)

[3. Test for differences across sex 7](#_Toc203923330)

[3.1 Hypothesis 7](#_Toc203923331)

[3.2 The test 7](#_Toc203923332)

[4. Test for differences across income groups 8](#_Toc203923333)

[4.1 Hypothesis 8](#_Toc203923334)

[4.2 The test 9](#_Toc203923335)

[5. Pearson correlation coefficients between LEF and variables 10](#_Toc203923336)

[5.1 Pearson’s correlation coefficient equation 10](#_Toc203923337)

[5.2 LEF and GDPPC correlation 10](#_Toc203923338)

[5.3 LEF and WA correlation 11](#_Toc203923339)

[5.4 LEF and HS correlation 12](#_Toc203923340)

[6. Pearson correlation coefficients between GDPPC and variables 13](#_Toc203923341)

[6.1 GDPPC and WA correlation 13](#_Toc203923342)

[6.2 Hypothesis 14](#_Toc203923343)

[6.3 The test 14](#_Toc203923344)

[6.4 GDPPC and WA correlation 15](#_Toc203923345)

[6.5 Hypothesis 16](#_Toc203923346)

[6.6 The test 16](#_Toc203923347)

[7. Regression model estimation (1) 17](#_Toc203923348)

[7.1 Regression Line 18](#_Toc203923349)

[8. Interpretation and statistical significance of R2 18](#_Toc203923350)

[8.1 Hypothesis 19](#_Toc203923351)

[8.2 The test 19](#_Toc203923352)

[9. Regression model estimation (2) 20](#_Toc203923353)

[9.1 Regression Line 20](#_Toc203923354)

[9.2 Hypothesis 21](#_Toc203923355)

[9.3 The test 21](#_Toc203923356)

[9.4 Which model is better? 22](#_Toc203923357)

[10. Prediction 23](#_Toc203923358)

[11. Conclusion 23](#_Toc203923359)

[12. References 24](#_Toc203923360)

# Introduction

Life expectancy at birth is to be considered the most important measure of health due to its significant relationship with other health indicators and ease of calculation (Khalsa, 2011). Additionally, life expectancy coincides with the measure of human development as an indicator of the human development index (HDI). As a result, this project aims to assess each determinant and what affect it has on the life expectancy of women (LEF) and life expectancy of men (LEM) of lower-middle and low income countries.

The Gross Domestic Product Per Capita (GDPPC) measures a country’s average income per person in $. It was seen that by fostering greater economic growth and development in a country, the GDP per capita raises the life expectancy at birth (Miladinov, 2020). This may be a result of people with higher incomes being able to afford basic human needs and therefore may have a negative effect on life expectancy for lower income countries.

Water Access (WA) represents the % of people using at least basic water services. According to Ritchie H and Roser M (2021), ‘the issue of unsafe sanitation… limited to low and lower-middle income countries…rates here are often greater than 50 deaths per 100,000’. This confirms a close relationship between clean water access and death rates.

Health Spending (HS) represents government expenditure per capita on health in $. A study carried out by Our World in Data depicts that ‘the population lives increasingly longer as health expenditure increases’ (Roser M, 2017).

I will be assessing the determinants against the dependent variables (LEF/LEM) using descriptive statistics, correlation, statistical significance tests and regressions.

# Descriptive Summary Statistics

## 2.1 The Female Life Expectancy (LEF)

Figure 1.1

|  |  |
| --- | --- |
| ***Life expectancy of Females*** | |
| Mean | 68.47810305 |
| Median | 69.73500061 |
| Mode | N/A |
| Minimum | 55.875 |
| Maximum | 79.60900116 |

This shows that throughout the 39 countries in the sample, the average female life expectancy is age 68, ranging between the ages of 55 and 79.

Figure 1.2

Figure 1.2 shows a near equal distribution, with the highest proportion of 28% belonging to the age range of 70-74 (most common) and the lowest being age range 60-64, with the top range showing 18% of countries have a high female life expectancy of 75-79.

## 2.2 The Male Life Expectancy (MEF)

Figure 1.3

|  |  |
| --- | --- |
| ***Life expectancy of Males*** | |
| Mean | 64.14571811 |
| Median | 65.46700287 |
| Mode | #N/A |
| Minimum | 51.45399857 |
| Maximum | 75.77500153 |

This shows that throughout the 39 countries in the sample, the average male life expectancy is age 64, ranging between the ages of 51 and 75. A slightly lower average and tighter range in comparison to females, suggesting that on average women tend to live longer than men.

Figure 1.4

Figure 1.3 shows a more unequal distribution, with the highest proportion of 36% belonging to the age range of 65-69 (most common) and the lowest being age range 75-79 (2%) which is the highest male life expectancy range.

## 2.3 Gross Domestic Product Per Capita (GDPPC)

Figure 1.5

Figure 1.5 presents a large percentage of countries (around 49%) to have a GDPPC between $1001-$2000. This could be a result of the given sample taken from low and low middle income countries. According to OECD (2014), evidence proves addressing inequality promotes sustained growth which suggests the low income countries given in the sample have a low GDPPC due to the link between income inequality and economic growth.

## 

## 2.4 Water Access (WA)

Figure 1.6

Figure 1.6 shows the highest proportion of countries being within the 91-100% water access range which means the majority of countries within the sample have access to basic water needs. Nonetheless, there are a proportion of countries (3%) who have only 31-40% of people with access to clean water which indicates poverty and therefore may have adverse effects on life expectancy.

## 2.5 Health Spending (HS)

Figure 1.7

Figure 1.7 shows the distribution of health spending to be skewed to the left between $0-$40. This suggests that the countries given within the two lowest income classes lack the finance to allocate towards health spending, moreover across the sample the highest spending doesn’t exceed $250 per person.

# Test for differences across sex

## 3.1 Hypothesis

To test whether there’s a difference between the average life expectancy of men and women, I will be conducting a hypothesis test with the following hypothesis:

H0 : = 0 (there is no difference between the average LEM and LEF)

H1 : 0 (there is a difference between the average LEM and LEF)

## 3.2 The test

Figure 2.1

|  |  |  |
| --- | --- | --- |
|  | LEF | LEM |
| Mean Life Expectancy () | 68.478 | 64.1457 |
| Standard Deviation (*S*) | 6.865 | 6.316 |

Equation 1 ()

* = 68.478
* = 64.1457

Equation 2 (*S*)

* = 6.865
* = 6.316

In Figure 2.1, ‘Equation 1’ calculates the mean life expectancy and ‘Equation 2’ calculates the standard deviation (by taking the square root of the variance) for the variables.

I will be testing the hypothesis at a 5% significance level and due to this being a two tailed test, where the test statistic can be either above or below zero, this would equal 2.5% for both tails. Referring to the standard normal distribution table as the sample size is >25, this returns a value of:

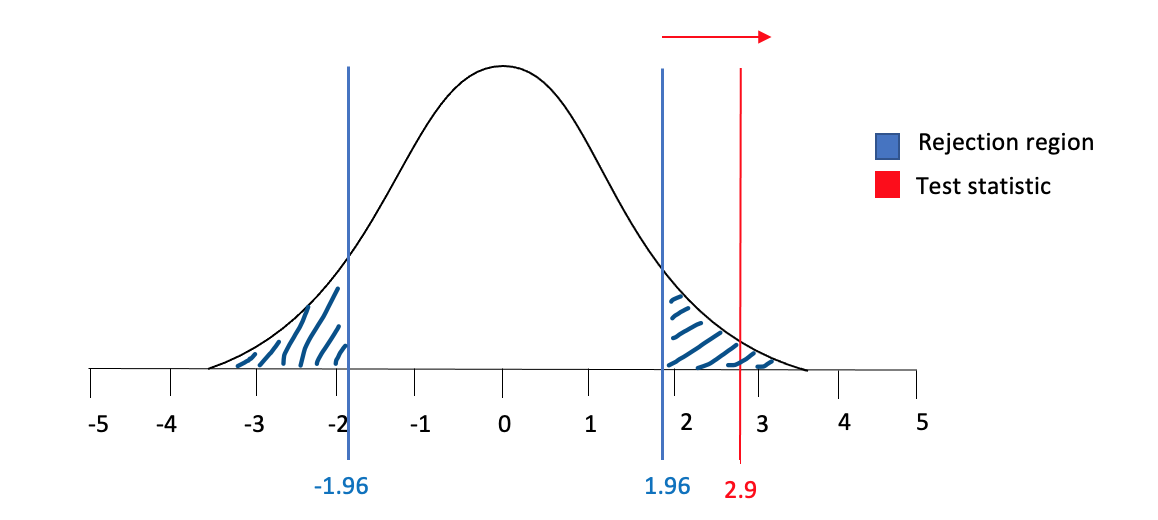
* Z0.025 = 1.96 (critical value)

Equation 3 (z)

* = 2.9 (test statistic)

Equation 3 is used to calculate the difference in the means as we have two independent samples from two populations.

Figure 2.2

(\*not to scale)

Z-test

We can reject the null hypothesis at the 5% significance level as the test statistic is within the rejection region in Figure 2.2, meaning we can be 95% certain that there is a difference between the average life expectancy of men and woman in the sample.

# Test for differences across income groups

## 4.1 Hypothesis

To test whether average female life expectancy is higher in ‘lower middle income countries’ than in ‘low income countries’, I will be conducting a hypothesis test with the following hypothesis:

H0 :

(average LEF in lower middle income is less than or equal to the average LEF in low income countries )

H1 :

(average LEF in lower middle higher than average LEF low income countries)

## 4.2 The test

Figure 3.1

|  |  |  |
| --- | --- | --- |
|  | Average LEF in lower middle income countries | Average LEF in low income countries |
| Mean Life Expectancy () | 70.0587 | 62.35 |
| Sample variance (s2) | 38.939 | 35.053 |

Equation 1

* = 70.0587
* = 62.35

Equation 4 (*s2*)

* = = 6.24
* = = 5.92

Due to having a small sample we use this formula for

* = = 6.1809

I will be testing the hypothesis at a 1% significance level and due to this being a one tailed test (as we only want to know if the average LEF in lower middle income countries is higher) this will equal 0.001 which I looked up on the t-distribution table:

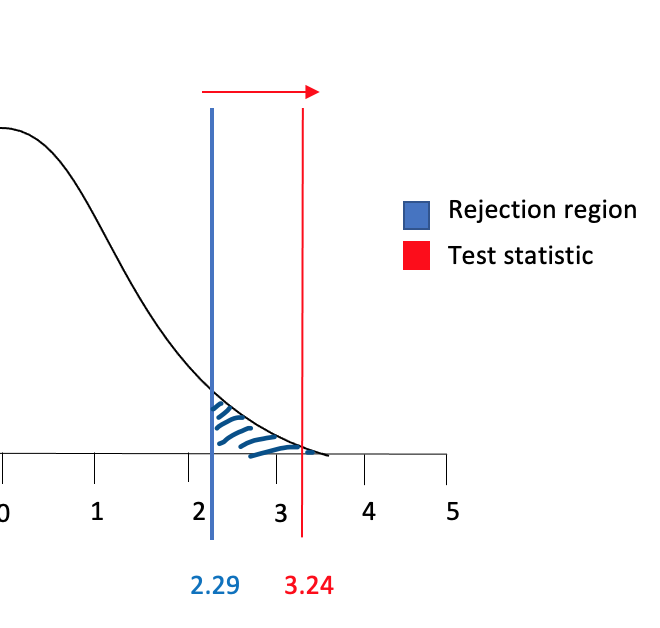
* t0.01 = +2.423(critical value)

Equation 3 (z)

* = 3.14 (test statistic)

Figure 3.2

(\*not to scale)



T-test

3.14

We can reject the null hypothesis at the 1% significance level as the test statistic is within the rejection region in Figure 2.2, meaning we can be 99% certain that the average LEF in lower middle is higher than average LEF in low income countries

# Pearson correlation coefficients between LEF and variables

## 5.1 Pearson’s correlation coefficient equation

Equation 6 (r)

## 5.2 LEF and GDPPC correlation

* = 0.6125

Figure 4.1

The correlation between LEF and GDPPC can be seen in figure 4.1. Equation 6 represents the strength of the linear association between the variables. As it’s above zero we can conclude there’s a positive corelation. Judging by the graph, it seems to be relatively strong however, a few countries with a GDPPC below 2000 have a higher female life expectancy than expected for a perfect correlation. Yet, correlation doesn’t imply causality and therefore we cannot be sure that a higher gross domestic product would equal a higher female life expectancy but the two can be closely linked whereby in lower income countries, people may not have the money for a high quality of life.

## 5.3 LEF and WA correlation

* = 0.7804

Figure 4.2

As expressed in equation 6, the correlation between LEF and WA shows a strong, positive correlation with the coefficient of 0.78 being close to 1 (perfect correlation). The graph highlights most points being close to the line. Not implying causality, it can be predicted that countries with a higher % of people accessing clean water would have a higher life expectancy due to water being a basic human need. Although, it can be seen that a country with near 80% WA has a LEF of 56, reinforcing the idea of considering other factors due to the sample being taken from selected low income and low middle income countries.

## 5.4 LEF and HS correlation

* = 0.6411

Figure 4.3

The correlation of LEF and HS from equation 6 shows a 0.64 correlation coefficient which is positive and somewhat strong. It shows that a higher amount spent on health is closely linked to a higher female life expectancy. Nonetheless, a cluster of data points below the line suggests health spending doesn’t have a strong relationship to the life expectancy age in comparison to other variables.

# Pearson correlation coefficients between GDPPC and variables

## 6.1 GDPPC and WA correlation

* = 0.6286

Figure 5.1

The correlation coefficient between GDPPC and WA is 0.6286 and figure 5.1 represents their relatively strong, positive relationship. It shows that a higher GDPPC is closely linked to a higher % of WA for countries.

## 6.2 Hypothesis

To test the statistical significance between GDP p.c. (GDPPC) and Water Access (WA) I will be using a hypothesis test to test that the Pearson Correlation Coefficient does have a significance.

H0 : = 0 (the coefficient is not statistically significant)

H1 : 0 (the coefficient is statistically significant)

## 6.3 The test

* = 39 – 2 = 37

I will be testing this hypothesis at the 5% significance level indicated by the alpha but as this is a two-tailed test (above or below zero), it will be 0.025. The coefficient is t-distributed and therefore requires ‘’ degrees of freedom. Using the t-table, we obtain a critical value of:

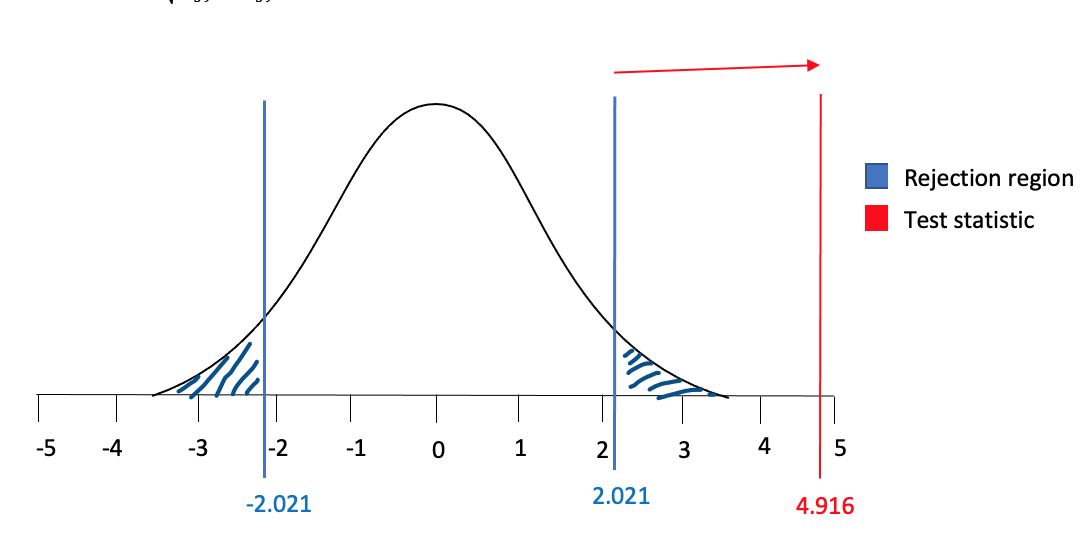
* *t*0.0025,37 = 2.021

Equation 7 (t)

* = 4.916

Figure 5.2

(\*not to scale)



T-test

Equation 7 represents the test statistic for this test thus returning a value is 4.916. We can reject the null hypothesis and be 95% sure that there is a statistically significant, positive correlation between GDP p.c. and Water Access.

## 6.4 GDPPC and WA correlation

* = 0.8158

Figure 5.3

The correlation between GDPPC and HS is high with a correlation coefficient of 0.81 which represents a strong, positive correlation. Figure 5.3 proves this idea with the data points clustered closely around the correlation line. It can be inferred that countries with a low GDPPC will have lower health spending. This could be due to there not being enough money for health spending to be considered a priority, or not enough money for any kind of spending in general.

## 6.5 Hypothesis

To test the statistical significance between GDP p.c. (GDPPC) and Health Spending (HS) I will be using a hypothesis test to test if Pearson Correlation Coefficient has a significance.

H0 : = 0 (the coefficient is not statistically significant)

H1 : 0 (the coefficient is statistically significant)

## 6.6 The test

* = 39 – 2 = 37

I will be testing this hypothesis at the 5% significance level for a two-tailed t -test thus, the critical value will be the same as in the previous test:

* *t*0.0025,37 = 2.021

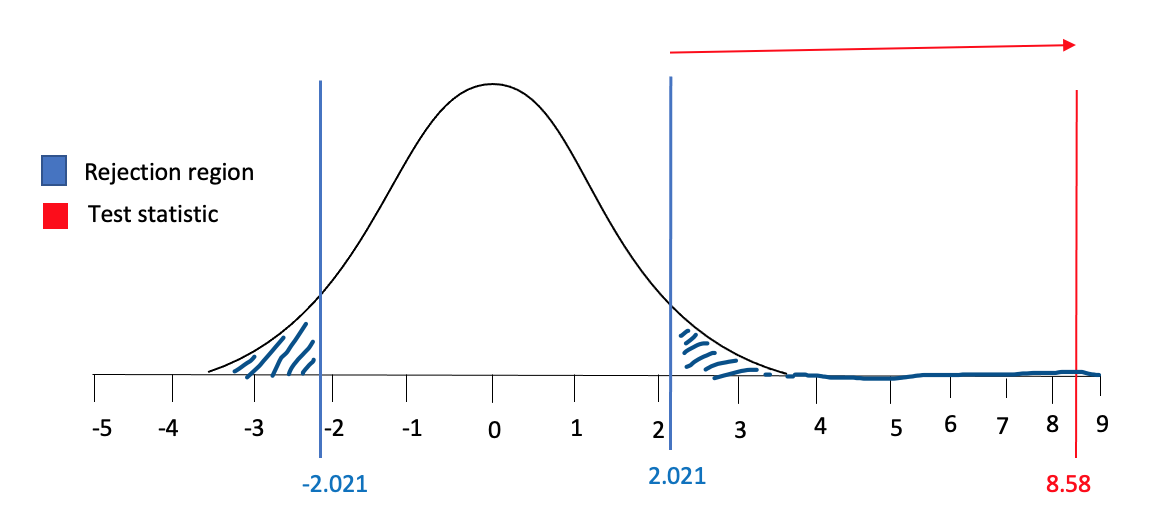
Equation 7 (t)

* = 8.58

Figure 5.4

(\*not to scale)

T-test

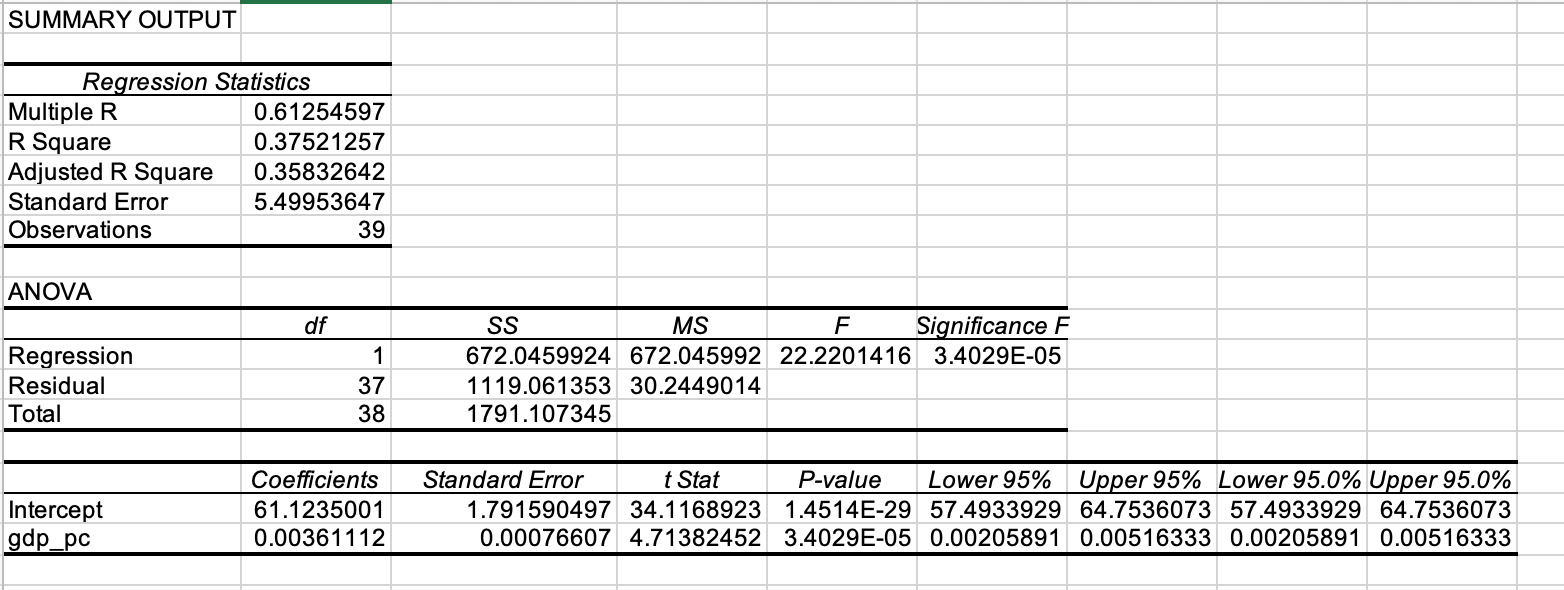


Equation 7 represents the test statistic for this test thus returning a value is 8.58. We can reject the null hypothesis and be 95% sure that there is a statistically significant, strong, positive correlation between GDP p.c. and Health Spending.

# Regression model estimation (1)

Whereby:

The LEF is the dependent variable which we assume will be explained by the X variable (GDPPC) with an intercept () slope () and error term ().

Figure 6.1

## 7.1 Regression Line

Equation 8 (slope coefficient, )

* = 0.00361112

Equation 9 (regression line)

* Intercept ‘ in the regression line equation in terms of ‘, represents that when holding GDPPC constant, we would expect the average female life expectancy to be 61.
* Intercept ‘ shows if GDPPC rises by $1, life expectancy would increase by 0.00361112. This value is low due to GDPPC usually being measured in $000’s, thus a $1000 increase in GDP would increase the LEF by 3.611 which would be more impactful.

# Interpretation and statistical significance of R2

Figure 6.1 shows the coefficient of determination (r2) to be:

R2 = 0.37521257

This value measures how much of the variation in female life expectancy is explained by GDP in the sample, in this case, it would be around 37.5%.

* = 0.37521257

## 8.1 Hypothesis

* H0 : 0
* H1 : 0

## 8.2 The test

This test follows an f-distribution and I will be testing at a 5% significance level ():

Looking this up on the F distribution table (upper 5% points), we return a critical value of:

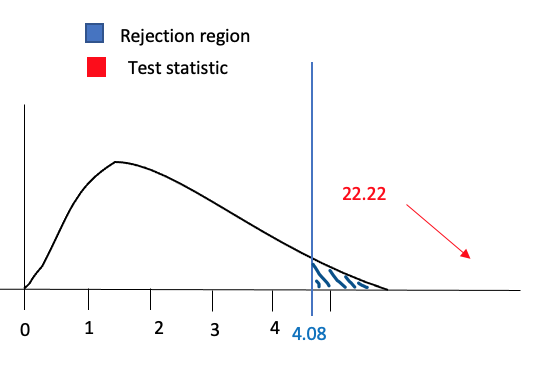
* =

Equation 10 (f-test)

* = 22.22

Figure 7.1

(\*not to scale)

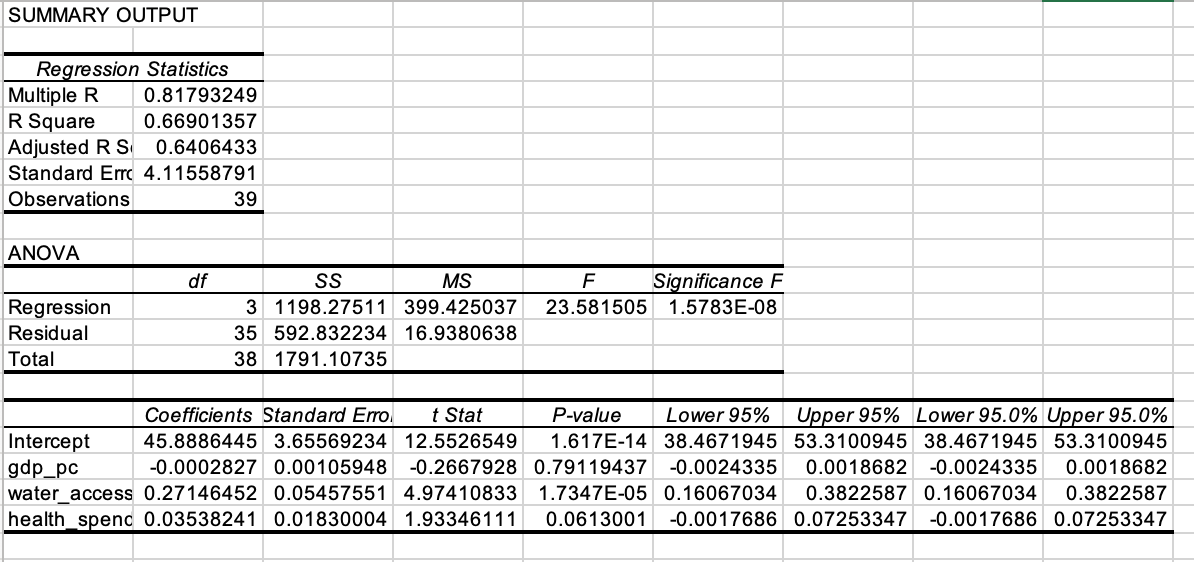


We can reject the null hypothesis and be 95% sure that the model has explanatory power as figure 7.1 shows the test statistic to be within the rejection region and far past the critical value of 4.08.

# Regression model estimation (2)

Whereby:

Figure 8.1



## 9.1 Regression Line

Equation 8 (slope coefficient, )

* = -0.00028266 (GDPPC)
* = 0.27146452 (WA)
* = 0.03538241 (HS)
* 45.8886445

Equation 9 (regression line)

* Intercept ‘ in the regression line equation in terms of ‘ represents that when holding GDPPC, WA and HS constant, we would expect the average female life expectancy to be 45.
* Intercept ‘ shows when holding everything else constant, if GDPPC rises by $1, the female life expectancy would decrease by 0.00028266 which implies GDPPC in this context has a negative effect on life expectancy. This wouldn’t be expected in the real world as if a country had a high GDPPC, we assume that they have more money for a better quality of life, thus impacting the life expectancy positively.
* Intercept ‘ shows when holding everything else constant, if the % of people with water access rises by 1%, the female life expectancy would increase by 0.27146452 which shows a significant positive impact
* Intercept ‘ shows when holding everything else constant, if health spending rises by $1, the female life expectancy would increase by 0.03538241 showing a smaller positive impact.

Figure 8.1 shows the coefficient of determination (r2) to be:

R2 = 0.66901357

## 9.2 Hypothesis

* H0 : 0
* H1 : 0

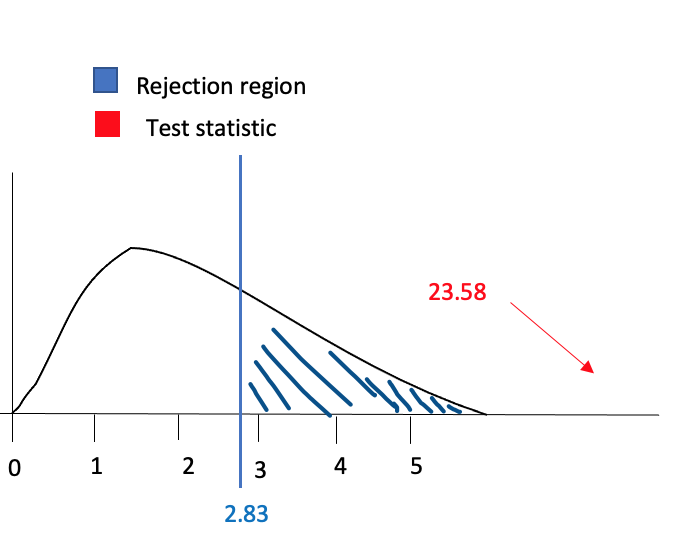
## 9.3 The test

This test follows an f-distribution and I will be testing at a 5% significance level ():

Looking this up on the F distribution table (upper 5% points), we return a critical value of:

* =
* = 23.58

Figure 8.2



We can reject the null hypothesis and be 95% sure that the model has explanatory power as figure 8.2 shows the test statistic to be within the rejection region and far past the critical value of 2.83.

## 9.4 Which model is better?

The second model is better because although the r2 may of increased due to the increase in variables, it explains more links between female life expectancy and differing variables. The first model may also be more susceptible to omitted errors due to there being a higher chance of having a value in the error term, such as health spending or an outside variable like population size. These may be correlated to GDPPC and thus the coefficient stated would be inaccurate.

# Prediction

* LEF prediction = 67.2955642 = 67

The female life expectancy prediction for a country with a $1500 GDPPC, 70% WA and $80 HS is age 67 which is just below the mean for the sample (68), suggesting this country may be within the low middle income group.

# Conclusion

Upon my analysis, each variable had an impact on life expectancy. The most valuable being

‘Water Access’. This is prevalent from the Pearson correlation between LEF and WA having the strongest coefficient and therefore the strongest relationship (figure 4.2). The multiple regression in 8.1 also shows water access to have the highest effect on the LEF in comparison with the other variables. GDPPC, although highlighting a strong positive correlation with LEF, shows positive and negative impacts in the regressions. Therefore, it is difficult to be 100% sure on the impact GDPPC has on life expectancy. Lastly, the effects of health spending on life expectancy remains constantly positive, nonetheless, the relationship is deemed to be the weakest of the three and so not as important when discussing life expectancy. Overall, due to the sample size being only taken from 39 low and low middle income countries, a possible improvement would be to include all countries that fit under this umbrella making it a population rather than the sample.

# References

*Inequality hurts economic growth, finds OECD Research* (2014) *OECD*. Available at: https://www.oecd.org/newsroom/inequality-hurts-economic-growth.htm(Accessed: December 8, 2022)

Khalsa, Dharma (2011). Life Expectancy: The Most Important Measure of Health. Alzheimer's Research & Prevention.

Miladinov, G. (2020) *Socioeconomic development and life expectancy relationship: Evidence from the EU accession candidate countries - genus*, *SpringerOpen*. Springer International Publishing. Available at: https://genus.springeropen.com/articles/10.1186/s41118-019-0071-0#citeas (Accessed: December 8, 2022).

Ritchie, H. and Roser, M. (2021) *Clean water*, *Our World in Data*. Available at:https://ourworldindata.org/water-access (Accessed: December 8, 2022).

Roser, M. (2017) *Link between health spending and life expectancy: The US is an outlier*, *Our World in Data*. Available at: https://ourworldindata.org/the-link-between-life-expectancy-and-health-spending-us-focus (Accessed: December 8, 2022).